EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2024

Turning optimal clasifiers into anomaly detectors

Motivation

- The most powerful architectures for supervised classification learn the physical information more efficiently.
- But… **how can we turn them into anomaly detectors and how good are they?**

Strategy

- Adaptation of 2-3 different classifier architectures with 3 methods to detect anomalies.
- No network optimisation (or minimal) was performed to avoid biases.
- Taking the average of scores from different hyperparameter choices.

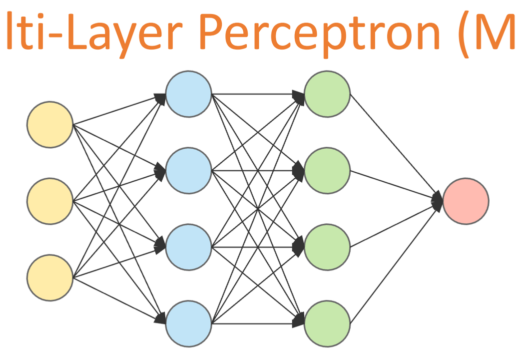
DarkMachines dataset

- Open data: dataset from *anomaly score challenge* [1].
- Event generation: *proton-proton* collisions at 13 TeV with Madgraph+Pythia.
- Detector simulation: simplified card for ATLAS detector at CERN with Delphes 3.
- Reconstructed particles (objects): jets, b-tagged jets, charged leptons, photons.
- Low level variables: object type, the four-momentum of the objects and the missing transverse momentum of the event.

Architectures

Networks implementation taken from [2]

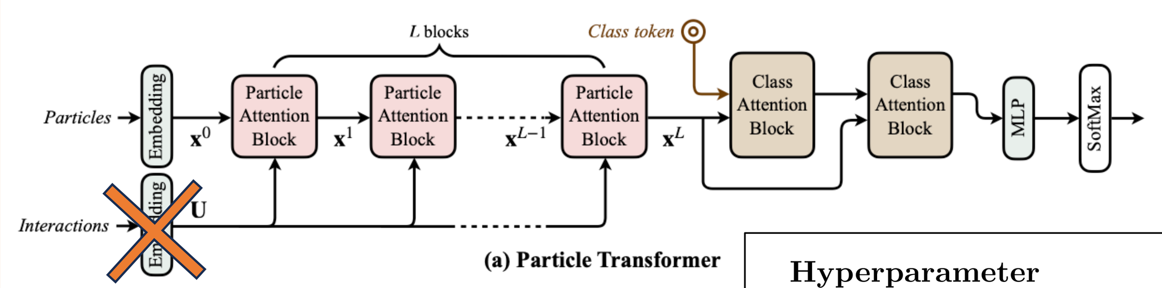
Multi-Layer Perceptron (MLP)



MLP

Dense(size = 16, activation = relu)
BatchNormalization
Dense(size = 8, activation = relu)
BatchNormalization

Particle Transformer (ParT)



No pairwise interactions

Hyperparameter	Value
Embed MLP dimensions	[128, 512, 128]
Pair embed MLP dimensions	[64, 64, 64]
Number of attention heads	8
Number of layers	8
Number of class attention blocks	2
FC dimensions	[64, 256, 64]
Auxiliary FC dimensions	[32, 32, 128]

ParT+ SM couplings

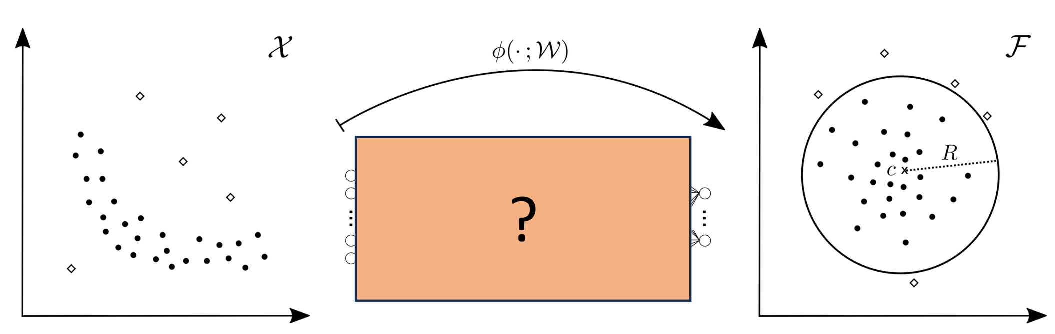
- Pairwise interactions
  - $\ln(m^2_{ij})$
  - $\ln(\Delta R_{ij})$
  - Physical information from the Standard Model: couplings.

Developed by this group

Techniques

Deep Support Vector Data Description (dSVDD) [3]

- Add an output layer with certain dimensions.
- Training: minimise distance to a centre in the hypersphere (anomaly score).
- Outliers are considered anomalies.
- Make ensemble [4] for different dimensions.

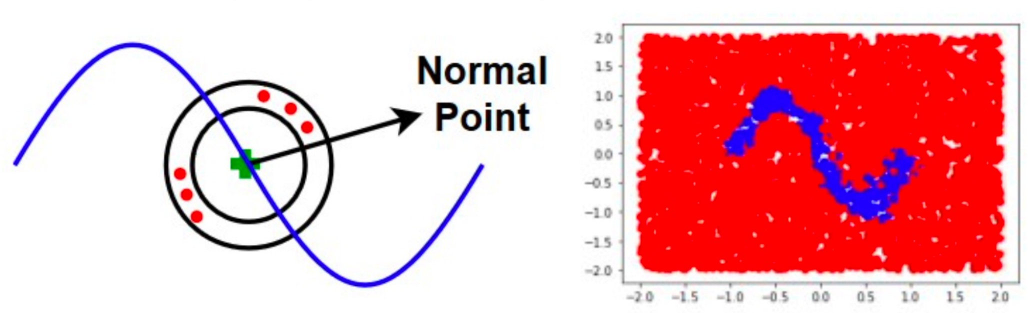


Deep Robust One-Class Classification (DROCC) [5]

- Background is assumed to lie in a low-dimensional manifold.
- Anomalous background events are generated and their location in the manifold is searched with an adversarial training.

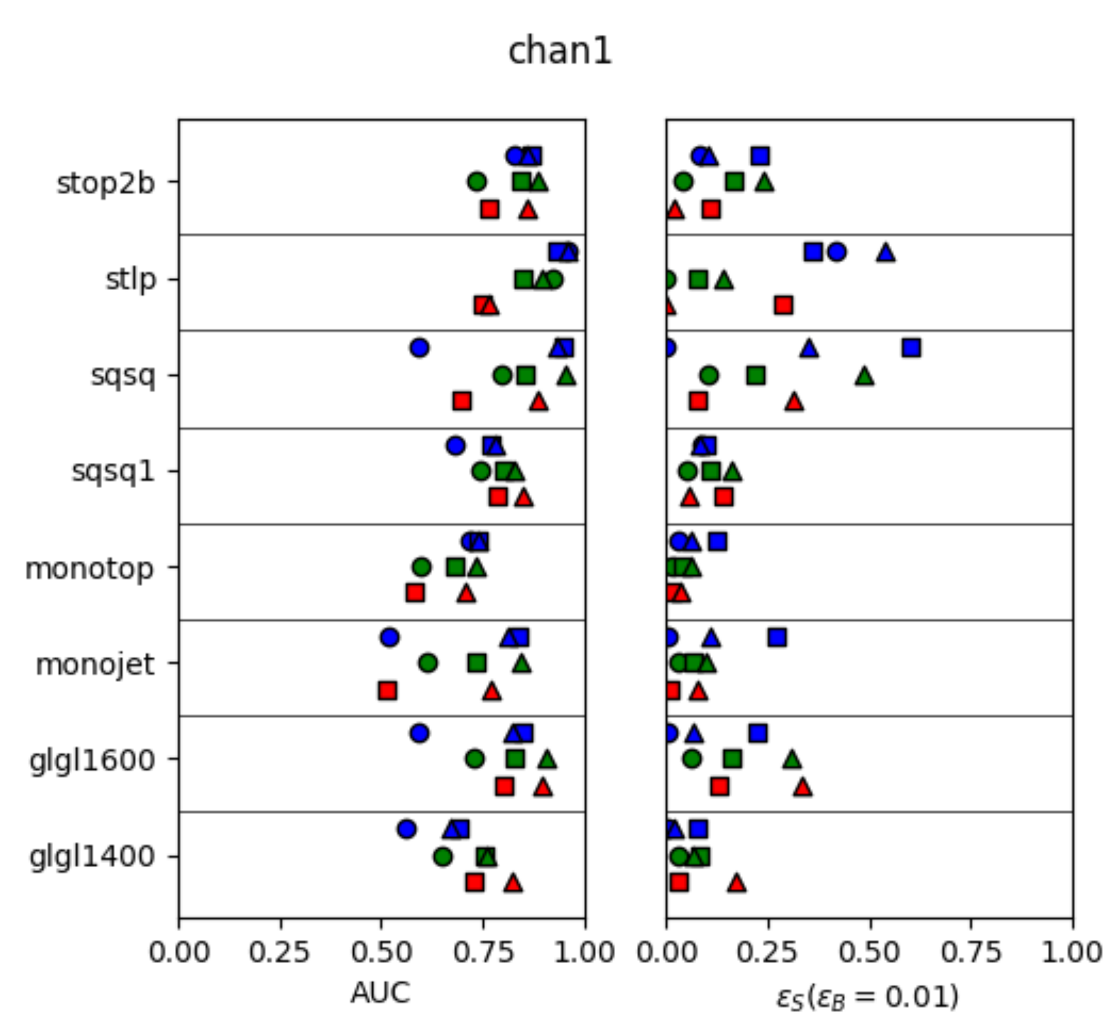
$$\sum_{i=1}^{n}[\ell(f_\theta(x_i),1) + \mu \max_{\tilde{x}_i \in N_i(r)} \ell(f_\theta(\tilde{x}_i),-1)]$$

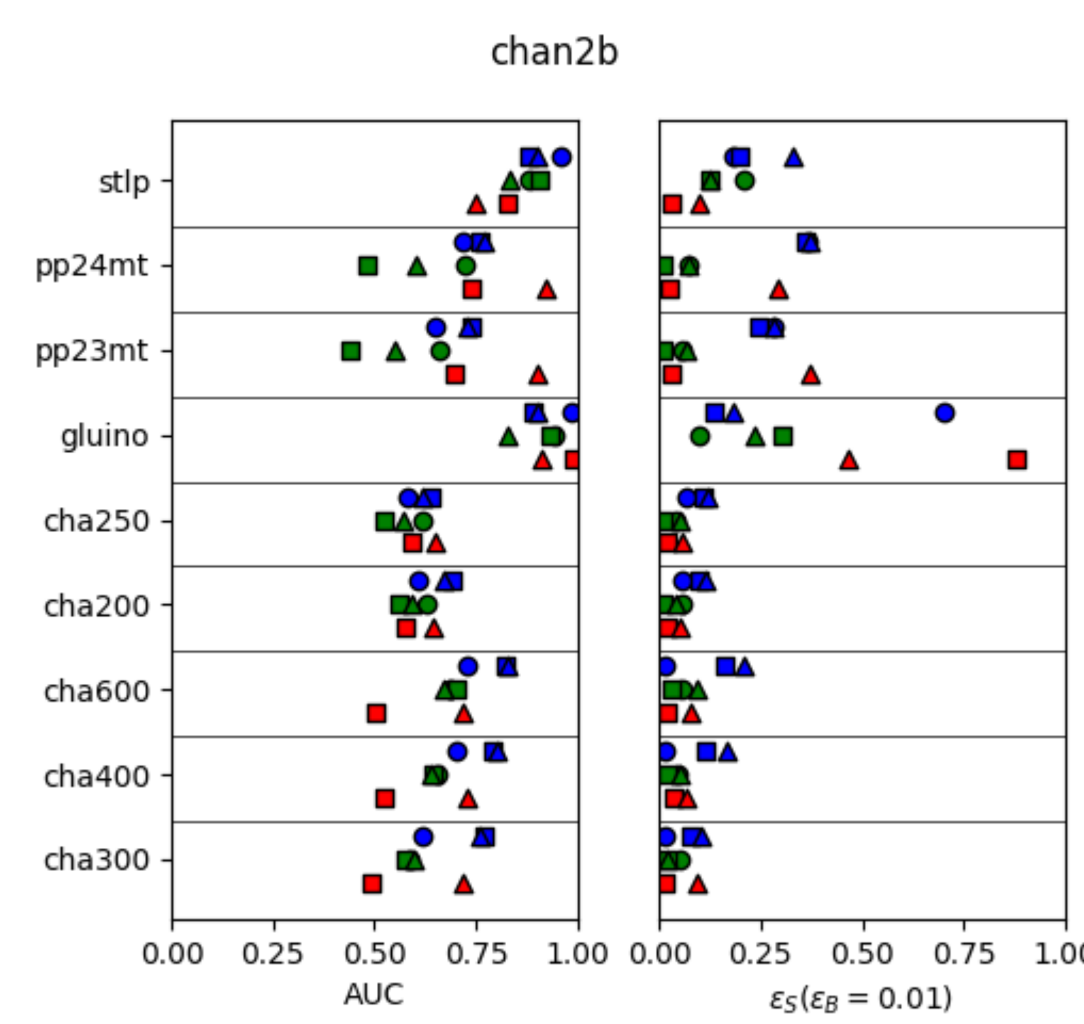
- Weakly supervised implementation.

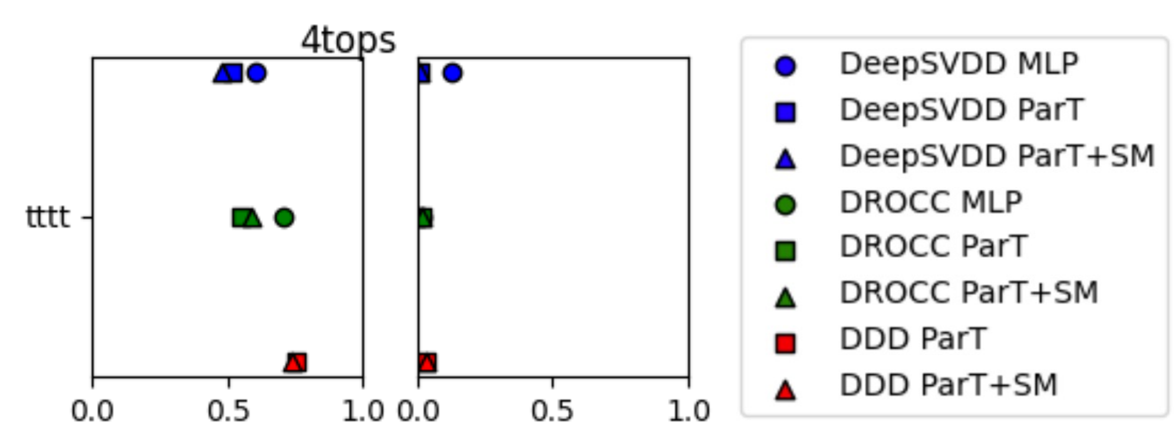


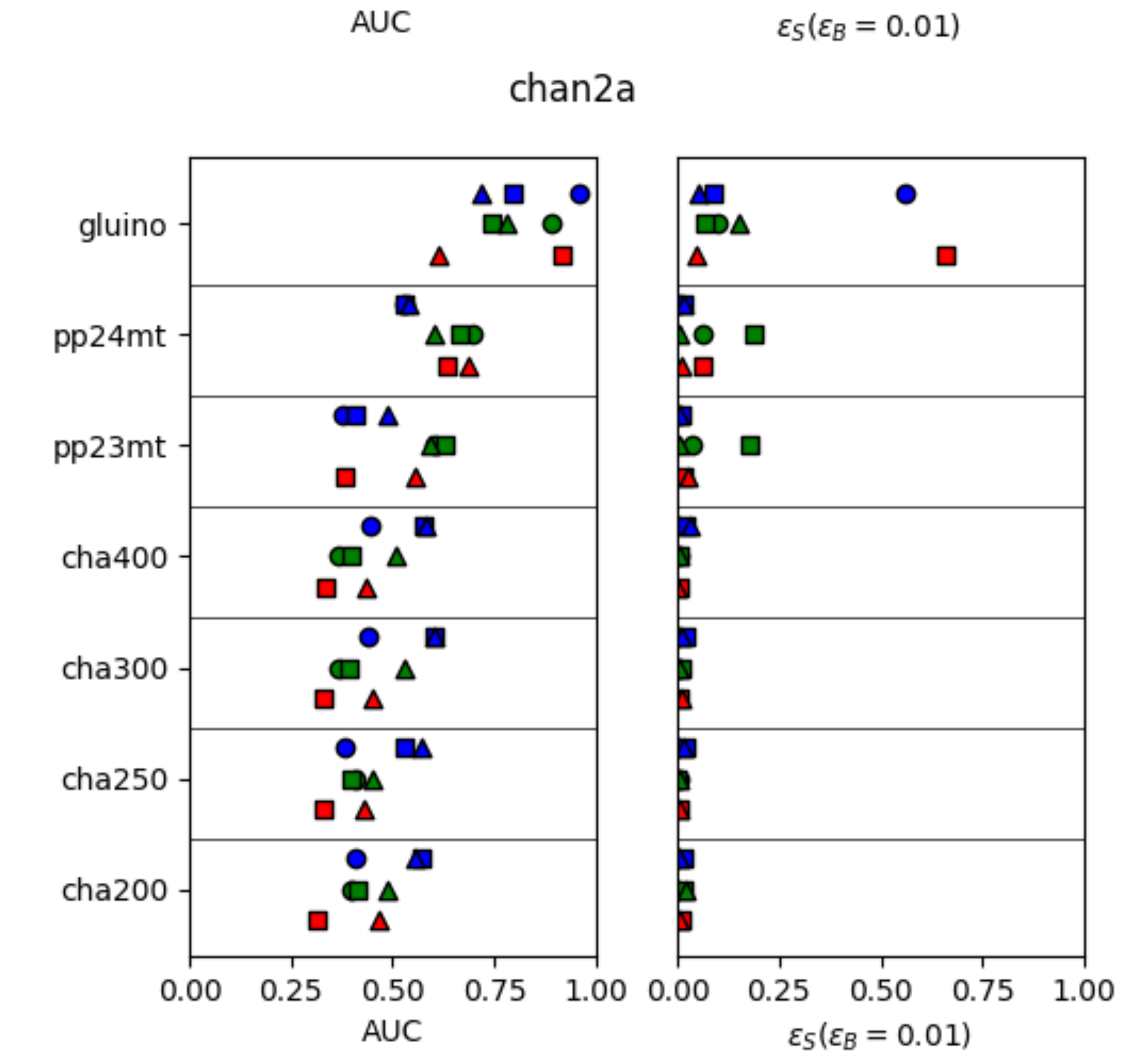
Discriminant distorsion detection (DDD)

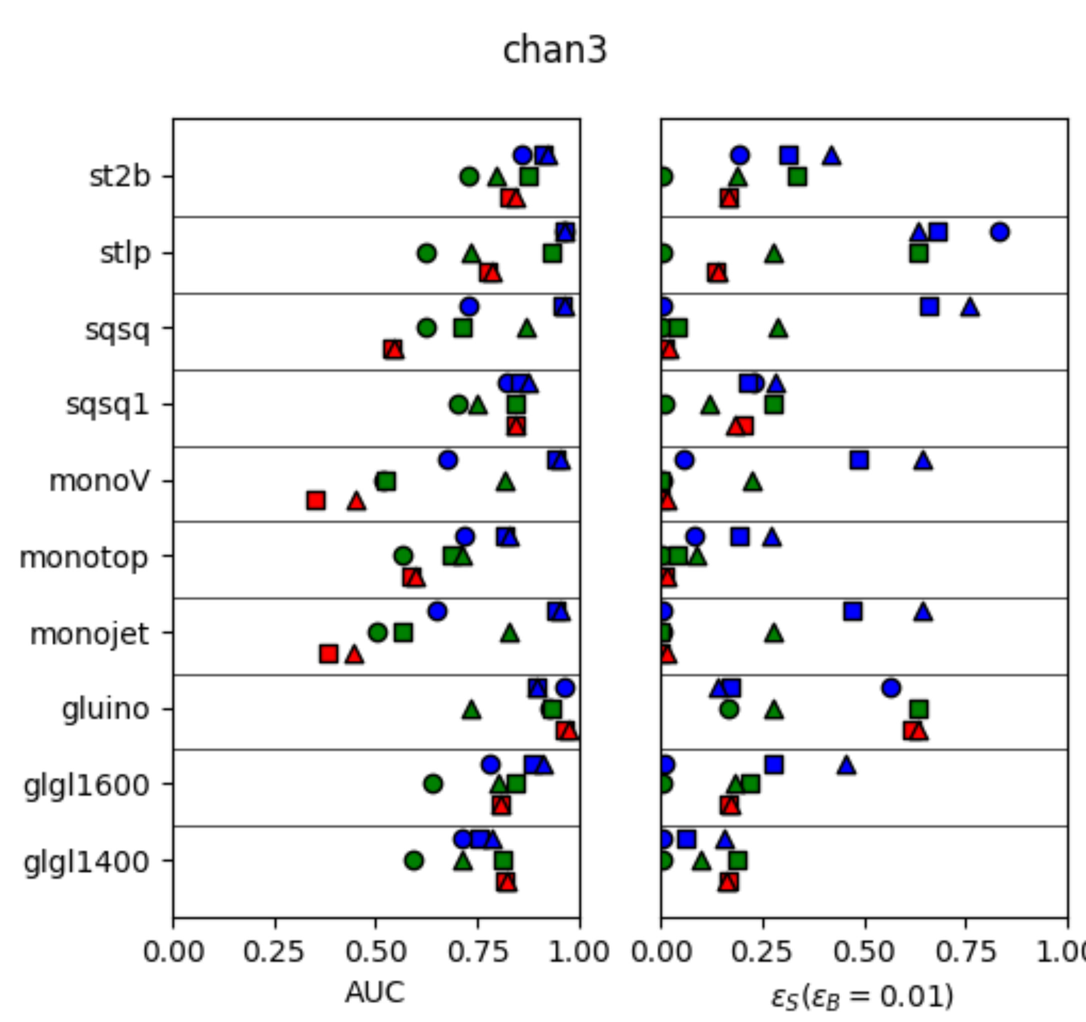
- New technique developed for this study.
- Anomalies look like distorted backgrounds.
- It creates a distorted training dataset:
  - It smeares the kinematic variables with a gaussian: scan on standard deviations.
  - Objects may be added or removed from each events: scan on probabilities.
- Training: discriminate *distorted bkg* vs *bkg*.
- Models with AUCs ~ 0.7-0.8 are picked up for testing on signals. Ensemble was made.











Conclusions

- Shown that we can take a supervised classifier and transform it into a (good) anomaly detector.
- **The best classifiers are -on average- better anomaly detectors:** ParT+SM in this case.
- Similar performances among the 3 methods. Compatible with anomaly score challenge results.
- A recommendation could be to use dSVDD and DDD in combination (fully unsupervised).
- The new method DDD discriminates between data with and without distortions. This opens interesting future research directions.
- A more detailed recipe will be found in the paper (very soon in arXiv).

[1] https://scipost.org/10.21468/SciPostPhys.12.1.043
[2] https://arxiv.org/abs/2211.05143
[3] http://proceedings.mlr.press/v80/ruff18a/ruff18a.pdf
[4] https://arxiv.org/pdf/2106.10164
[5] https://arxiv.org/pdf/2002.12718.pdf

Team members: Sascha Caron (Nikhef, Radboud U), José Enrique García Navarro (IFIC, CSIC-UV), María Moreno Llácer (IFIC, CSIC-UV), Polina Moskvitinaa (Nikhef, Radboud U), Mats Rovers (Radboud U., Nikhef), **Adrián Rubio Jiménez (IFIC, CSIC-UV)**, Roberto Ruiz de Austri (IFIC, CSIC-UV), Zhongyi Zhang (Bonn U.).
Email: Adrian.rubio.jimenez@cern.ch